

Weekly Report

November 12, 2017

1 Work

This week, we have completed k-means method to speed up the embedding part of LargeVis. However, there exist a bug in our code. The projection result is not satisfying. We plan to solve this bug next week. After fix this bug, we will take some time tuning parameter. In addition, thanks to Dongming, we are able to write and compile latex on web (like overleaf). I start to write our dimension reduction paper on <http://10.76.0.186:8083/>.

1.1 工作进度

Table 1: 工作进度

TASK	PROGRESS	DATE
dimension reduction	Add k-means method to speed up the embedding	11.30
*2Vec survey	Write a little framework	12.30

2 Paper Reading

2.1 Product quantization for nearest neighbor search

这个方法是通过子空间的聚类，对原始空间进行空间划分。在计算两点之间的距离的时候，采用空间到空间的近似距离来代替，从而提高计算效率。文章提出的方法学习了单词及其相关拟声词之间的关系，这是基于对同一段声音的不同标签学习了一个相似性。

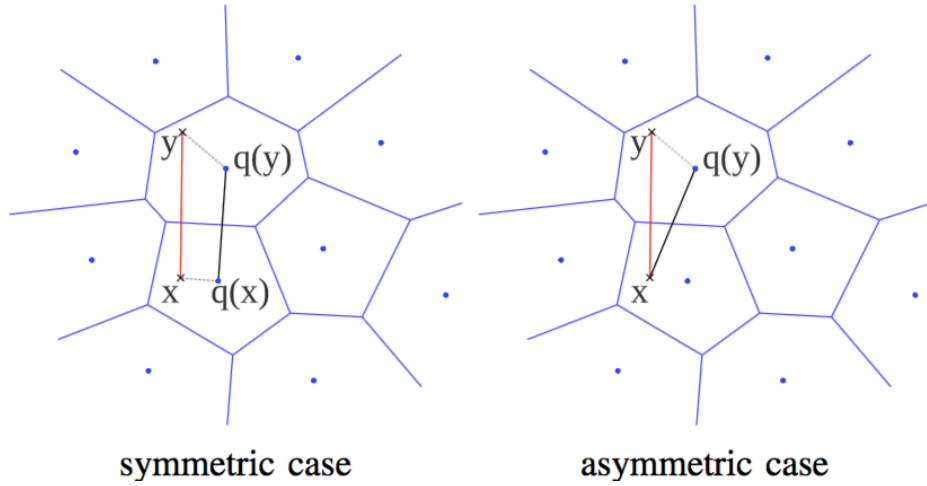


Figure 1: Product quantization

2.2 EdWordle: Consistency-preserving Word Cloud Editing

本文提出了一种对词云的编辑方法，当编辑完一个词之后，周围的词会基于局部的结构进行重新布局，最大限度保持局部结构。



Figure 2: EdWordle

2.3 Visualizing Big Data Outliers through Distributed Aggregation

一般在数据中的异常点，由于数量比较少，在聚合的过程中（往往是基于数值聚合）会被忽略掉。本文提出了一种类似于聚类的方法，在数据密集的地方有效降低数据的数量，同时保持数据稀疏区域的数据分布。同时，本文还提到，通过降维等方法观察高维数据往往会得到不正确的结果。在高维空间中异常的数据，在低维空间可能就是正常的。

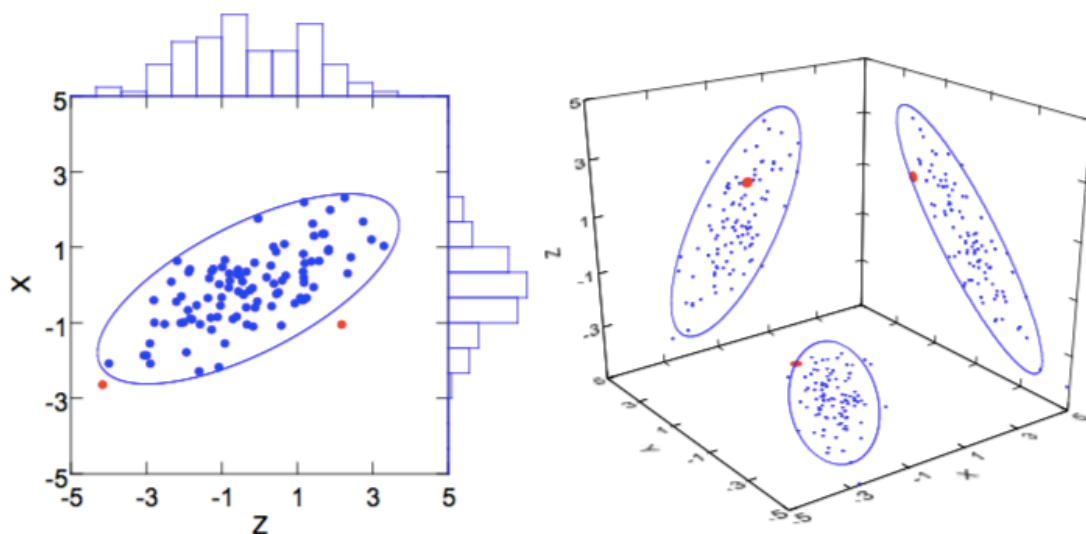


Figure 3: Outliers

2.4 Clustering Trajectories by Relevant Parts for Air Traffic Analysis.

本文提出了一个对于飞机轨迹聚类方法。

2.5 A graph optimization method for dimensionality reduction with pairwise constraints

提出了半监督的graph- optimized locality preserving projections (GoLPP)，主要是在边的权重上对于同一类节点权重设为1，不同类节点权重设为0，没有标记的节点权重按照距离定义。